DOCUMENT RESUME

ED 075 933 EA 005 152

AUTHOR Garner, William T.
TITLE The Identification of an Educational Production Function by Experimental Means.
PUB DATE 28 Feb 73
NOTE 48p.; Paper presented at American Educational Research Association Annual Meeting (58th, New Orleans, Louisiana, February 25-March 1, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Academic Achievement; Cost Effectiveness; Costs; Decision Making; Economic Research; Educational Strategies; Experiments; Grade 8; Input Output Analysis; Mathematics Instruction; *Performance Criteria; Performance Factors; Speeches; Statistical Analysis; *Statistical Studies; Tables (Data); Teaching Methods; *Time Factors (Learning)
IDENTIFIERS Cobb Douglas Function; *Education Production Functions; Mastery Learning

ABSTRACT

Eighth grade students, randomly assigned to three criterion performance levels, studied matrix algebra in three programed lessons. Student achievement, ability, time spent, and other measures were obtained; and minimal variance criterion performance levels, analogous to production isoquants, were attained. A Cobb-Douglas (log-linear) function was estimated by regression, with output (criterion levels) exogenous and time to mastery used as a dependent variable. The use of the function to predict the time required for various student ability and performance combinations is illustrated. Costs and implications for equity/efficiency decisions in school management and finance are discussed under various assumptions. A companion document is EA 005 176. (Author)

FILMED FROM BEST AVAILABLE COPY

ED 075933

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

THE IDENTIFICATION OF AN EDUCATIONAL PRODUCTION FUNCTION BY EXPERIMENTAL MEANS

William T. Garner

Northwestern University

Presented at the annual meeting of the
American Educational Research Association
February 28, 1973, New Orleans, La.

EA 005 152

I

INTRODUCTION

This paper[1] may be viewed from three major perspectives. It reports the identification of an educational production function by a method previously unused for that purpose, and in stricter conformity with economic theory than most earlier educational production function studies. It also deals with a relationship which has received too little attention in most studies of school performance or student achievement: namely, the relation of the allocation of time to student achievement. In addition, this paper may be viewed as a report on the use of a mastery learning strategy.

The motivation for each of these three components was described in a paper presented to this Association two years ago.[2] In the interest of brevity, I shall consider that earlier paper to be the introduction to the present one. Readers who desire a fuller explanation of the background and rationale of my work are urged to consult that earlier paper.

In brief, my argument is that most previous educational production function studies suffer from a number of shortcomings: 1) improper specification of outputs; 2) inattention to the distribution of outputs; 3) failure to use longitudinal data on changes in outputs (indeed, the use of cross-

---

1 What follows is a preliminary and summary description of a project to be reported in detail in The Identification of an Educational Production Function by Experimental Means, Ph.D. dissertation in progress, The Department of Education, University of Chicago.

2 "Inputs and Outputs in the Educational Process," a paper by this author presented at a symposium on "What Can Economics Contribute to Education?," AERA, New York, February 4, 1971. Available from the author.

sectional rather than longitudinal data on "value-added" means that many so-called "production-function" studies do not qualify on this ground alone); 4) improper specification of inputs; 5) the assumption that an average of school inputs is applied to each student; 6) the use of school, or even district-wide, averages rather than data for individual students; 7) the assumption that reported stocks of inputs are fully utilized; 8) the assumption that inputs are utilized efficiently; and 9) aggregation over diverse instructional technologies.

In view of these many shortcomings, it is not surprising that educational "production-function" studies have not led to more powerful or more useful results. Take, for example, the assumption that the average amount of each school input is applied to each student. To the extent that this represents actual educational practice, it is an important reason to expect background variables to account for more of the variance in achievement than school variables, except when schools show large variance in input means. And to the extent it does not reflect actual practice, it, of course, weakens the models, particularly with respect to the application of results to improve within-school allocations. It must be kept in mind that production (or, education) is a process which takes place over time; it is not an instantaneous event. Thus if we are to know how much of a particular resource is used by students, we must have observations on the allocations of student time with those resources.

In a very real sense, it is the student, rather than the school, who "produces" his own gains in achievement. Each student has his own array of talents and skills which he combines with available school resources in order to "learn," i.e. to produce increments in his measured achieve-

ment. Thus again there is a need for data on the resources of individual students--their general ability and their prior achievement in specific areas--and on the time allocated by students to learning tasks.

Finally, we must know the distributions of inputs and outputs, and the relation between them. With such knowledge, we can answer the question "Given a particular time allocation of school and student resources, what outputs can be expected?" Or, to pose the question asked in this research, "Given particular distributions of desired outputs, what time allocations of school and student resources are required to produce them?"

The information required for rigorous educational production studies cannot be easily obtained by observation of natural school environments or from conventional school records. Hence, I advocate an experimental approach designed specifically to elicit such production information. It would undoubtedly be wasteful to attempt to investigate the production relationships in all educational situations by experimental means. But study of some key educational processes, such as early reading or arithmetic mastery, could yield information of importance. A sufficient number of such experimental studies might provide a basis not only for improving the efficiency of educational production, but for rationalizing the distribution of achievement, especially achievement of the basic learning skills upon which later learning both in school and on the job are presumably founded. Such experiments must include:

1) Clearly defined outputs and inputs,

2) Specification of the technology, and

3) Assurance of technical efficiency.

Condition one, with the present state of educational measurement, will limit choice of outputs to a fairly narrow range of cognitive skills. This limitation, however, is a practical, not a conceptual one. Condition two must be understood in its broad sense, in which technology is not merely hardware of various kinds, but organizational and instructional procedures as well. Condition three is met largely through exercising care in the conduct of experiments, and is one of the reasons for preferring experimental to naturalistic settings. The present study attempts to meet all of these requirements.

This Study in Brief

The general approach was to give three brief programmed instruction lessons on "matrix arithmetic" (matrix algebra) to about 100 eighth grade public school students in two racially and socio-economically homogeneous schools. In addition to each student's pretest score on the specific subject matter, his or her prior achievement scores on a standard test in reading and arithmetic were obtained from school records. Students were randomly assigned to one of nine possible treatment sequences; each treatment consisting of a required score on lesson tests. A record was kept by the student of time spent on each part of the sequence. The resulting data was subjected to regression analysis using a model assuming multiplicative interaction among the independent variables, and using elapsed time as dependent variable.

Two basic elements of the technology employed were programmed instruction format and a "mastery learning" strategy. The reader is assumed

to be familiar with programmed instruction. The essential features of mastery learning[1] are:

1) Stated behavioral objectives. These are predetermined things students are intended to learn, that is, learning tasks which are to be mastered.

2) Formative tests of each student's relevant prior achievement before instruction begins, and of his progress toward the current objectives. Such tests are for informational purposes, not for assigning grades.

3) Diagnosis of the deficiencies of each student with respect to the objective behaviors, based on the formative tests.

4) Instructional sessions, which may include group and individual instruction.

5) Individual (sometimes paired or group) review and remedial work, based on the diagnoses.

6) Criterion performance levels. These are defined as proportions correct on tests of the objective behaviors.

By using the mastery learning approach, it was possible to arrange and define the conditions of learning so that most students mastered specified skills at given levels of performance, a model which closely parallels the economic one of a firm which can produce a uniform product at various levels

---

1 See James H. Block (ed.), Mastery Learning, Theory and Practice, (New York: Holt, Rinehart, and Winston, Inc., 1971.

of output. In production terms, the model employed here could be specified in part as follows:

1) The uniform kind of output produced is mastery by all (or nearly all) students of the defined skills (objectives) of a particular course or subject area.

2) The level of output is the criterion performance level, as defined above.

3) The technology of production is programmed instruction, mastery learning, and other conditions and procedures detailed below.

The next section of this paper describes the administrative and data-collection procedures of the experiment. Selected results will be presented in Section III. The results will be discussed and the paper concluded in Section IV.

II

PROCEDURES

This section describes the subjects and materials employed in this experiment, and summarizes the experimental and data-collection procedures. Thus it is in large part a specification of the relevant technology.

Teaching Materials

Certain general considerations guided the search for teaching materials appropriate for this study. First, a subject-matter was sought which would permit the clear, or relatively clear, statement of learning outcomes ("output") in measurement terms. This consideration made mathematics a natural candidate for use here. Second, a programmed format was desired, in order to render the presentation of material to students as uniform as possible. The object was to reduce teacher influences upon student learning, as no characterization of "key" teacher attributes was to be attempted here. Third, the materials should lend themselves to the manipulation of student achievement levels, a necessary condition for this experiment but one foreign to standard curriculum materials.

These considerations were largely satisfied by the matrix arithmetic lessons developed by James H. Block for his dissertation research.[1] Block's

---

[1] James H. Block, "The Effects of Various Levels of Performance on Selected Cognitive, Affective, and Time Variables," unpublished Ph.D. dissertation, The University of Chicago, 1970.

general approach, described in detail in his thesis, was to adapt the introductory lessons of a matrix algebra textbook to eighth-grade level. The principal problem was not to make the material easier to learn, but rather, more difficult! The original textbook, being in programmed format, was designed to provide "errorless" learning, through the provision of frequent practice and repetition in the text. The efficacy of the unmodified text is shown in Block's report that, in a sample of "above average" eighth-graders, student performance on unit tests averaged 80 percent correct.[1] By eliminating redundancies and making other changes, Block obtained a set of three lessons such that "[t]he mean performance of students using the modified textbook, wherein both practice and repetition were minimal, was only 50 percent."[2]

For each lesson, Block constructed a twenty-item test of lesson content. Two additional equivalent verisions of each lesson test were then generated according to procedures described by Bormuth.[3] Item-specific review and practice material was then constructed, so that a person missing any given test item could be directed to review content specific to that item. Thus the complete array of available curricular material for each lesson consisted of:

---

1 Ibid., p. 48.

2 Ibid., p. 50

3 John R. Bormuth, On the Theory of Achievement Test Items, (Chicaog: University of Chicago Press, 1970).

1) An initial "text" or lesson;

2) A formative test;

3) Review material;

4) Two additional versions of the formative test.

The reader should now see how these materials allow manipulation of student achievement, that is, student mastery of lesson content as measured by scores on lesson test. The typical student using the modified text can be expected to score perhaps 50 percent on the initial formative test of twenty items. It is assumed that the student has already mastered the content of the ten items he answered correctly. Suppose we desire the student to improve his mastery to 80 percent, or 16 items correct. The student can be given review material specific to six missed items. He is not assigned more than six items to review since we do not (for purposes of the experiment) desire him to achieve more than 80 percent mastery. After review, he is tested again. If he still falls short of mastery, say by one item, he is given new review materials and subsequently another test. This sequence was not continued indefinitely in this experiment; results by the third formative test were accepted as final for the particular lesson. How well this system was able to produce mastery at target levels will be reported in Section III.

Subjects

An attempt was made to locate a population of about 100 eighth-graders of one race and with relatively homogeneous socio-economic characteristics.[1]

---

[1] As originally conceived, two separate experiments were planned using student (subject) groups of markedly contrasting socio-economic backgrounds for purposes of comparison. Unfortunately, time and money limitations made only one experiment possible.

A school district organized on a "neighborhood school" basis in a residential suburb west of Chicago cooperated, but it was not possible to arrange the study in just one school (due to the small size of the schools). Nor could the two schools in the "most similar" neighborhoods be included in the study. Thus, two schools in somewhat different parts of the community were obtained. The resulting divergence in student socio-economic backgrounds is illustrated in Table 1, comparing student family size and parents' level of education.

Table 1

MEANS OF STUDENT FAMILY CHARACTERISTICS, BY SCHOOL

	(1) No. Persons in Household	(2) No. of Children Under 18	(3) Father's Educ. (Years)	(4) Mother's Educ. (Years)
School #1	5.6	3.1	15.6	14.7
School #2	5.7	3.2	13.1	12.8
T Value*	.271	.284	4.27**	3.58**

* T Value for difference of means test
** Significant at better than .01 level (two tailed)

Students in the two schools come from families which do not differ importantly in total size (Table 1, Column 1), or in the number of children under eighteen years old (Column 2). The educational attainment of fathers and mothers, however, does differ "significantly" between the

schools (Columns 3 and 4). The detectable effects of these between school differences in parents' level of education will be discussed in the following section.

A total of 135 students, comprising all the eighth-graders in the two schools, participated in the experiment, but due to absences or incomplete records, only 110 students are included in the analysis.

All subjects are caucasian.

School Administered Standard Tests

All subjects (for which complete records are available) took the Stanford Achievement Test, Form W, Advanced Battery, during October 1970. Each student's raw scores on the Paragraph Meaning and Arithmetic Concepts subsections were obtained from school records. The SAT tests are designed with "generous" time-limits, in order to be "fundamentally power tests and not speed tests."[1]

Experimental Design

The teaching materials comprised three sequential lessons in matrix arithmetic. The object was to have each student learn each lesson, but not with the same level of mastery for all students. Thus, by a fundamentally arbitrary decision, three performance, or mastery, levels were selected as targets for each lesson. These were the 65,80, and 95 percent levels,

---

[1] SAT Directions for Administering, p. 3.

which correspond to 13, 16, and 19 items correct on a 20-item test. The students of each school were separately ranked according to SAT Arithmetic Concept raw score, then taken sequentially in runs of three and randomly assigned to one of three groups. The groups were then randomly designated as 65, 80, and 95 percent mastery groups. These assignments were recorded as each student's target mastery level for lessons one and two.[1] The randomization was then repeated to assign students to target performance levels for lesson three. As a result of this procedure, some students (about 1/3) received the same performance level assignment for all three lessons, while some received sequences which raised or lowered the target performance level after the second lesson.

This random reassignment of performance level permits the investigation of the effect of performance level sequence on learning.

The experimental design may be illustrated as a nine-cell matrix, as in Table 2. The number in each cell indicates the number of students with the corresponding assignments.[2] (Ignore the parenthetical T_1 for the present.) The main diagonal of the matrix represents uniform performance level

---

1 Block found that, on the average, students took about as long to complete both lessons one and two as to complete lesson three alone. Thus it was decided here to assign the same target mastery level to lessons one and two for each student. Some simplification of this kind was required by the limits of the sample size.

2 N=107. Complete data is available for 110 subjects, but three students followed incorrect sequences during the experiment due to experimenter error.

assignments for all three lessons. Cells above the diagonal are assignments in which the performance level increases from the second to the third lesson; in cells below the diagonal, the sequence decreases.

Table 2

PERFORMANCE LEVEL ASSIGNMENT MATRIX

Lessons One & Two Performance Level Assignment	Lesson Three Performance Level Assignment: 65%	80%	95%
65%	(T1) 14	(T2) 12	(T3) 11
80%	(T4) 11	(T5) 10	(T6) 12
95%	(T7) 14	(T8) 12	(T9) 11

The design establishes conditions in which students of varying initial resource (prior achievement) combinations produce different levels (i.e. performance levels) of a uniform kind of output (matrix algebra skills). The range of observed resource/output combinations is undoubtedly greater than would ordinarily be observed in non-experimental conditions. The experiment is analogous to much agricultural research, or to "response surface" methodology more generally, although a larger number of cells and observations would improve the analogy.

Administration

In November, 1971, the experiment was conducted sequentially in the two

schools, with the assistance of the regular classroom teachers and a graduate student. Two weeks were required at each location. No alterations in school schedules were made for the experiment. Indeed, the somewhat complicated class schedules in both schools provided less than ideal experimental conditions in terms of continuity and record keeping. All student-kept records were checked frequently for completeness and accuracy. On the whole, the students cooperated extremely well and proved themselves excellent record-keepers.

Summary

A total effective sample of 110 eigth-graders was obtained in two schools rather less homogeneous in population than hoped for. Instructional materials developed by James Block for a similar design but different purpose were adapted for use here. Students were randomly assigned to one of three "mastery" or performance levels (defined as percentage correct on a twenty item test) for the first two, and again for a third lesson. Student self-paced instruction of the programmed-format lessons was administered by the experimenter and assistants, during separate two-week periods at each school.

The data available for analysis may be grouped into five categories:

1) Student family background information as reported by student answers to questionnaire;

2) Student SAT Arithmetic Concepts and Paragraph Meaning raw scores;

3) Student performance level assignments;

4) Student pretest, formative tests, and post-test scores;

5) Student time (self-recorded) on all study and test activities.

The analysis of these data is taken up in the following section.

III

R E S U L T S

The first question to be taken up concerns the degree to which the ideal of the experimental design was fulfilled. Then the results of several alternate multiple regressions with student time as the dependent variable will be presented and interpreted. Finally, certain subsidiary matters, such as the effect of performance level sequence on rate of learning, will be discussed.

Student Mastery

Students were randomly assigned to one of three performance levels for each lesson. But what was the actual or realized performance of the students? The experimental problem was to raise the performance of students whose initial formative test scores were below the assigned level, and to intercept students with satisfactory scores, so that they would not exceed their assigned level. In operational terms, the ideal desired results were performance level scores with means of 13, 16, and 19 respectively, and with zero variance.

I was unable to obtain zero variance about the target scores, although in fact almost one-third (30.9%) of the students did attain exactly their pre-assigned performance levels on all three lessons. But I expect the reader to allow that zero variance, while ideal for theoretical reasons, is perhaps too rigorous a standard to fulfill in practice. I offer in its place the concept of "minimal relative variance,"

by which I mean that the mastery procedure is effective when the variance around student target scores is small compared to the variance in the ability and prior achievement scores of the students.

A comparison employing this minimal relative variance concept is made in Table 3, in which--for each lesson and target performance level--the coefficient of variation of student final scores on each lesson is compared to that for student ability and prior achievement scores in each group.

(INSERT TABLE 3)

Table 3

COMPARISON OF THE RELATIVE VARIATION IN STUDENT SAT, PRETEST, AND EXPERIMENTAL LESSON SCORES, FOR EACH LESSON AND PERFORMANCE LEVEL ASSIGNMENT

		(1) Variable	(2) Mean	(3) Standard Deviation	(4) Coefficient of Variation (3÷2)
LESSON ONE	PL13 (65%) Group	ACONRW	18.135	5.779	0.319
		PRESUM	1.243	1.038	0.835
		TOTSC1	14.378	2.190	0.152
	PL16 (80%) Group	ACONRW	18.647	6.134	0.329
		PRESUM	1.265	1.163	0.919
		TOTSC1	16.000	1.181	0.074
	PL19 (95%) Group	ACONRW	19.103	6.711	0.351
		PRESUM	1.333	1.364	1.023
		TOTSC1	18.308	1.575	0.086
LESSON TWO	PL13 (65%) Group	ACONRW	18.135	5.779	0.319
		PRESUM	1.243	1.038	0.835
		TOTSC2	13.595	2.254	0.166
	PL16 (80%) Group	ACONRW	18.758	6.165	0.329
		PRESUM	1.303	1.185	0.910
		TOTSC2	15.849	1.278	0.081
	PL19 (95%) Group	ACONRW	19.270	6.854	0.356
		PRESUM	1.324	1.396	1.054
		TOTSC2	18.595	0.832	0.045

(CONTINUED)

Table 3 (Continued)

	(1) Variable	(2) Mean	(3) Standard Deviation	(4) Coefficient of Variation (3÷2)
LESSON THREE				
PL13 (65%) Group	ACONRW	18.878	6.501	0.344
	PRESUM	1.293	1.123	0.869
	TOTSC3	14.293	1.692	0.118
PL16 (80%) Group	ACONRW	18.275	6.400	0.351
	PRESUM	1.171	1.150	0.982
	TOTSC3	16.114	1.183	0.073
PL19 (95%) Group	ACONRW	19.152	5.885	0.307
	PRESUM	1.455	1.325	0.911
	TOTSC3	18.758	0.708	0.038

Definitions: PL13, etc., is the assigned performance level; ACONRW is the SAT Arithmetic Concepts test raw score; PRESUM is the score on a matrix arithmetic pretest; TOTSC1, etc., is the actual final or "total score" attained on each lesson.

Note that in Table 3, Column 2, the mean "total score" (TOTSC) in each lesson should be, ideally, either 13, 16, or 19, according to the performance level assignment. In fact, the mean for the lower performance level tends to be nearer 14, that for the middle level is about on target, and that for the highest level tends to be about 18.5. Thus the procedures employed appear to be somewhat less effective at intercepting learning at the lower performance level than in raising it to the higher criterion levels.

Column 4 gives the relative variation of the student scores in each group, for each lesson. Here we see that the coefficient of variation of scores on the experimental lessons is typically less than that of the students' SAT raw scores by a factor of 4 or 5, and less than the variation in pretest scores by an even larger factor (except, again, for the lowest performance level).

I would ask the reader to mentally contrast such low-variance outcomes with conventional classroom procedure, in which typically a relatively wide--rather than narrow--range of outcomes is tolerated. Indeed, a wide range of outcomes is perhaps encouraged by some teachers, as an aid to "making a curve" for grading purposes.

In summary, the instructional strategy employed for this experiment did not produce "ideal" results, but did in general obtain student performance narrowly concentrated near target scores, especially for the two higher performance levels. These student outcomes show far less vari-

ation[1] than the pre-experimental measures of relevant student resources, and support the view that under appropriate conditions similar levels of mastery can be attained by students of widely varying initial abilities. Furthermore, these minimal variance outcomes near predetermined (i.e., exogenous) performance levels seem to reasonably approximate an economic model of production at various levels of a uniform kind of output.

Selection of a Regression Model

For the reasons outlined in my earlier paper, the simple linear additive regression model was rejected in favor of a curvilinear model, the so-called Cobb-Douglas function which, however, is linear in logarithms and so may conveniently be estimated by ordinary least squares methods.

A single instructional technology was employed for the experiment; no comparison of technologies was attempted. Student "entry" or prior resource levels, as measured by SAT and pretest scores, were not manipulable for experimental purposes (apart from random assignment of students). Student target performance levels were manipulated, with considerable success as shown above. Thus the principal experimentally observed variable was student time spent at each stage. This situation is opposed to typical school practice, where student time spent on a topic tends to be bounded by the teacher's lesson plans or the pace of

---

1 One caveat to be noted is that the coefficient of variation may be biased if a test is subject to a "ceiling effect." This problem and other assessments of the effectiveness of the mastery learning procedures employed here are discussed in the detailed report of this research (see note 1, page 1).

the class as a whole, and student performance is a variable. Here, student achievement was exogenously determined, and student time was allowed to vary. Therefore, the basic function to be estimated was:

$$Y = a + b_i X_i + b_j X_j + u \qquad (10)$$

where: $Y = \log_e$ time in minutes

$a = \log_e$ constant term

$X_i = \log_e$ of a vector of student test scores

$X_j = \log_e$ of a vector of treatment (student assignment) dummies

$u = \log_e$ residual.

Transformations of Selected Variables

Student SAT Arithmetic Concepts and Paragraph Meaning raw scores were converted into ratios according to the formula $r = (p / 1 - p)$, where p is the proportion the raw score bears to the total points possible on the test. The logarithm of r was then taken. This monotonic transformation, known as the "logit" transformation, was used to stretch out the upper end of the distribution and obtain a better conditioned matrix for regression procedures. The logit values thus obtained were used directly in the regressions.

The student pretest variable (PRESUM) was also transformed slightly, according to the formula: NEWPRE = (PRESUM + 1). This was necessitated by the presence of a pretest score of zero for some students, since the logarithm of zero is undefined. The natural logarithms of resulting NEWPRE scores were used in regressions.

Dummy Variables

A number of "dummy variables" were constructed in order to include certain qualitative states in the regression analysis.[1] In general, N "states" or "qualities" may be represented by (N - 1) dummy variables, coded 1 or 0 to represent the presence or absence of the quality. For example, this experiment involves nine treatment groups. (Recall that students were assigned to one of three target performance levels for a pair of lessons, then were assigned to one of three levels for a third lesson. Thus a 3 x 3 assignment matrix results, containing 9 [9 = 3 x 3] cells.) Applying the rule for construction of dummy variables, these nine cells can be represented by eight (8 = 9 - 1) dummies, coded 1 if the student is a member of the cell, 0 otherwise. In this fashion, the students of all but one cell are identified by a "1" signifying membership in a particular cell. Students in the remaining cell receive a "0" for all eight of the dummy variables, signifying membership in none of the other eight cells. This cell becomes the reference group against which membership in the other cells is measured. Thus the regression coefficient for each cell or treatment dummy represents the net effect of being in that cell compared to the reference group. The particular cell to be used as the reference group is an arbitrary decision, and may be

---

1 An explanation of the use of dummy variables may be found in Jacob Cohen, "Multiple Regression as a General Data-Analytic System," Psychological Bulletin, Vol. 70, No. 6 (1968), 426-43, reprinted in Bernhardt Lieberman, Contemporary Problems in Statistics (New York: Oxford University Press, 1971), 421-440.

selected for analytical convenience.

The use of dummy variables in this way can be shown to be equivalent to analysis of variance procedures, but the multiple regression context has the added advantage of including quantitative variables as well.[1]

Regression Results

We have at last surmounted the preliminary obstacles (I hope to the reader's satisfaction), and can now proceed to identify a production function as promised in the title of this essay. We shall take a direct route to the more interesting results.[2]

We naturally expect students with higher levels of relevant prior achievement to spend less time on new learning, other things equal. This expectation is borne out in Regression #1, as shown by the negative slope coefficients for variables three, four, and five. Thus, as the student's Arithmetic Concepts, Paragraph Meaning, or pretest score is higher, the student will spend less time for a given amount of new learning, and conversly. It should be remembered that the regression equation coefficients are for logarithms of the variables; interpretation will be discussed below.

Variables 6 through 13 are treatment dummy variables, corresponding to the treatment matrix cells shown in Table 2 (see page 13), as T1, T2, etc.

---

1 Ibid.

2 Additional results, tests, and refinements will be included in the detailed report.

Treatment 5, the 80% performance level assignment for all three lessons, was selected as the reference group in construction of dummy variables. As shown in Regression #1, the t-statistics for the individual treatment variables are generally not high, with the exception of that for T9, which is significant at the .01 level. Taken as a group, however, the treatment dummy variables do contribute significantly to the total variance explained. This is shown in the following manner: Regression #2 is similar to Regression #1, below, but omits the treatment dummy variables. As the treatment assignments are independent of the variables in Regression #2, an F-statistic may be computed to estimate the significance of the additional variance explained by the addition of treatment variables in Regression #1, as compared to Regression #2. The additional variance explained is almost 10%, and the F-statistic for this increment is 2.49 (8,95 d.f.), which is significant at $p < .05$.

(INSERT REGRESSIONS #1 & #2)

REGRESSION #1

	Dependent Var.	Independent Variables						
	(1)	(2)	(3)	(4)	(5)	(6)		
Var. Name	LNTAT	CONSTANT	LGTACN	LGTPAR	LNPRE	T1		
Coefficient		5.023	-0.1975	-0.1982	-0.1156	-0.1058		
t-value		60.63	-3.86 ***	-3.60 ***	-2.27 *	-1.00		
		(7)	(8)	(9)	(10)	(11)	(12)	(13)
Var. Name		T2	T3	T4	T6	T7	T8	T9
Coefficient		-0.0968	0.0165	-0.1060	0.1308	0.0597	0.0816	0.2706
t-value		-0.84	0.15	-0.96	1.19	0.57	0.75	2.42 **

R^2 0.57

F-statistic 13.03 (10,96)

Num. obs. 107

* $p < .05$
** $p < .01$
*** $p < .001$

Definitions: LNTAT = Log_e of total time

LGTACN = Logit of SAT Arithmetic Concepts raw score

LGTPAR = Logit of SAT Paragraph Meaning raw score

LNPRE = Log_e of matrix arithmetic pretest

T_i = performance level assignment (treatment) dummies

REGRESSION #2

	Dependent Var.	Independent Variables			
Var. Name	LNTAT	CONSTANT	LNPRE	LGTACN	LGTPAR
Coefficient		5.048	-0.117	-0.195	-0.186
t-value		103.3	-2.18	-3.69	-3.29
Var. Name					
Coefficient					
t-value					

R^2 0.48

F-statistic 31.88 (3,103)

Num. obs. 107

Definitions: LNTAT = Log_e of total time

LNPRE = Log_e of matrix arithmetic pretest

LGTACN = Logit of SAT Arithmetic Concepts raw score

LGTPAR = Logit of SAT Paragraph Meaning raw score

Further explanation of the variance in total time spent by students is obtained by adding dummy variables to represent school and sex effects. Sex differences in school learning behavior are familiar, with the general finding that at least through grade school (including eighth grade) girls tend to receive higher marks than boys, to be disciplined less often, and so on. Thus to capture any systematic differences associated with sex, a single dummy variable (SEX) was coded 1 for girls and 0 for boys. If, as might be expected, girls use less study and test time than boys, the resulting regression coefficient will have a negative sign.

A dichotomous variable (SCHL) was created to capture any effects associated with the two schools. Presumably any systematic differences between the schools (if uncorrelated with other variables in the study) will be captured by this variable, although it might be impossible to specify exactly the source of the difference within the school settings. Thus such a variable is not analogous to an experimental "treatment," as no specific treatment components are identified. Instead, it is analogous to a machine or plant dummy in a regression analysis of production: a host of unspecified differences may be grouped by some readily identifiable feature, in this case by school.

Regression #3 is similar to Regression #1, but adds the school and sex dummy variables. The R^2 has now risen to 0.62, an increase of .05. This gain has a significant F-statistic of 6.25 (2,93 d.f.). The other variables in the regression retain relatively their same coefficients and t-values. SEX has a statistically significant slope coefficient of nearly the same magnitude as the Paragraph Meaning coefficient. Its negative sign indicates that girls tend to take less time to criterion than boys, other things equal.

(INSERT REGRESSION #3)

REGRESSION #3

	Dependent Var.	Independent Variables						
Var. Name	LNTAT	CONSTANT	LNPRE	LGTACN	LGTPAR	SEX	SCHL	T1
Coefficient		4.9611	-0.1041	-0.2319	-0.1866	-0.1419	0.0872	-0.0447
t-value		68.88	-2.13	-4.65	-3.48	-2.71**	1.68	-0.51
Var. Name		T2	T3	T4	T6	T7	T8	T9
Coefficient		0.1177	0.0877	-0.0434	0.1643	0.1150	0.1058	0.3286
t-value		0.44	0.93	-0.46	1.80	1.32	1.14	3.47

R^2 0.62

F-statistic 11.83 (13,93) d.f.

Num. obs. 107

** $p < .01$

Additional Definitions: SEX: 1 = girl

0 = boy

SCHL: 1 = School #1

0 = School #2

The SCHL variable coefficient in this regression does not reach conventional levels of statistical significance, but is large enough to suggest that systematic differences between the groups may be correlated with their school of attendance. (In other regressions to be reported, SCHL does enter with statistical significance.) There are a number of possible such differences. Different average mathematical and reading achievement, as indicated by SAT scores, is an obvious candidate. These scores in turn may be associated with different family incomes and preferences in the two school attendance areas--differences which affect student access to learning resources at home and motivation and performance in school.

Another difference between the groups may be a "teacher effect." All students in the experiment had mathematics with their same teacher the previous school year: each school has only one teacher for seventh and eighth grade mathematics. Thus teacher influence on student work habits or study skills (or on unmeasured mathematical knowledge) may account for "school" differences.

A final source to be mentioned of possible difference between schools is that "the" experiment itself may actually be two experiments. The schools were visited sequentially, and some changes in the experimenter's behavior, "learning" how to administer the experiment, if you will, may have influenced outcomes. Every attempt was made to make administration uniform between schools, but discrepancies, no doubt, crept in.

A large number of other variables were available for inclusion in regression analyses. These include several family environment variables, coded for the most part as dummy variables. These were: DESK, indi-

cating presence or absence of a desk or table the student considers his own at home (a home resource for school-related behavior); FATHER, MOTHER, indicating presence or absence of each parent (natural or by adoption); READTO, whether or not the child was read to daily while of pre-school age, according to his recollection; MOTHASP, FATHASP, indicating the student's perception of whether or not the parent expects him to enter college; and TALK, whether or not student and parents frequently discuss his schoolwork. Also included were the logs of two numerical variables: LNFED, log of father's education in years; and LNEDSP, the log of student's expected schooling in years.

It was not anticipated that these variables, or many of them, would have strong impact in terms of regression analysis. Although a plausible argument can be made that most of them should have some bearing on student achievement or time to criterion, such influence might not be clearly revealed in an experiment of relatively short duration which departs from conventional school procedure and in which homework as such plays no part. In addition, much of this information was collected in the hope that the original project design (a sample including strongly contrasting socio-economic groups) would eventually be fulfilled. In the absence of a contrasting group, most of these variables show relatively little variance, and hence cannot have much explanatory power in the present regression.

The results of a regression including the earlier variables and those just discussed are presented in Regression #4. Among the new variables added, only the log of father's education in years approaches statistical significance. (Mother's education is highly correlated with

father's education, but is less statistically significant than father's education when substituted in the below and other regressions.) Inclusion of the eleven new variables does increase the R^2 slightly over that of Regression #3, but the F-statistic for the increase is only 0.64 (11,82 d.f.). Thus no importance can be attached to the small gain in variance explained.

(INSERT REGRESSION #4.)

REGRESSION #4

Dependent Variable: LNTAT

Independent Variables

V-name	Coef.	t-value	V-name	coef.	t-value
C	6.213	6.29	FATHER	0.029	0.42
LNPRE	-0.079	-1.49	MOTHER	0.107	0.61
LGTACN	-0.218	-3.96*	READTO	0.045	0.66
LGTPAR	-0.154	-2.61*	MOTHASP	0.101	1.13
T1	-0.040	-0.43	FATHASP	-0.017	-0.21
T2	0.123	0.42	TALK	0.009	0.13
T3	0.148	1.37	LNEDSP	-0.243	-0.69
T4	-0.023	-0.23			
T6	0.171	1.70			
T7	0.142	1.49			
T8	0.073	0.72			
T9	0.329	3.24			
SEX	-0.189	-3.01*			
SCHL	0.099	1.54			
SELFCON	-0.049	-0.789			
DESK	-0.050	-0.716			
LNFED	-0.285	-1.63			
LNFMSZ	-0.006	-0.064			

R^2 0.65

F-statistic 6.44 (24,82), $p < .001$.

Num. obs. 107

Definitions: See text

* sig. at $p < .02$

Summing up, we have seen that more than 60% of the variance in time spent by individual students can be explained with knowledge of achievement test scores, sex, school, and performance level assignments.

The regressions reported above show, for a given technology, the relative influence of prescribed performance levels and varying student resources on total study and testing time spent by students. Because they embody this information, they may be interpreted as educational production functions. It might be more accurate to call them "time functions," since the dependent variable is elapsed time which, strictly speaking, is allocated or "spent" by students rather than "produced." The estimated equations probably should not be called "cost functions" because they include no factor prices, for the simple reason that none exist for such variables as student achievement test scores. Nevertheless, there is a fundamental duality between cost and production functions, and with knowledge of prices or appropriate assumptions about them, cost functions can readily be obtained from production functions.[1] As the equations estimated here reflect underlying technical production relationships in the setting described, they will be referred to "production functions." In particular, I nominate Regression #3 as "the" production function identified in this paper.

Interpretation of Regression Results

The use of elapsed time as the dependent variable here facilitates a

---

[1] R.W. Shephard, Cost and Production Functions, Princeton: Princeton University Press, 1953.

rather straightforward interpretation of the regression results. For discussion purposes, we select Regression #3 as containing the most readily interpretable and generally most significant variables. Then the predicted (log) time to completion for a student with any combination of characteristics (within the range observed in this study) is found by inserting the appropriate student scores into the equation estimated by Regression #1.

Suppose, for example, we want to estimate the required time for a "high-ability" student, Haynes, to complete the learning sequence with an 80% (middle level) performance requirement throughout. To be specific, it is assumed that Haynes' SAT raw scores are one standard deviation above the sample mean, he is male, attends the first of the two schools studied, and scored two correct answers on the matrix arithmetic pretest. The following are the values to be inserted in the equation:

1. LNPRE = 1.098 (Natural log of pretest score transformed by adding 1; i.e. in this case log 3)
2. LGTACN = 0.51 (Logit of Arithmetic Concepts raw score of 25; approx. equal to mean plus one s.d.)
3. LGTPAR = 0.93 (Logit of Paragraph Meaning raw score of 43; approx. equal to mean plus one s.d.)
4. SEX = 0 (Males are coded zero.)
5. SCHL = 1.0 (The first school studied is coded 1.)
6. T_i = 0 through 13. (The reference group assignment was chosen here, hence all treatment dummies are coded zero.

These scores are multiplied by their respective (estimated) coefficients, and added together with the constant term to give the estimated

log time for this student. The resulting term is exponentiated to give time in minutes. These operations are carried out in Table 4, and give as estimated time of 104 minutes for Haynes.

(INSERT TABLE 4)

Table 4

METHOD OF CALCULATING ESTIMATED TIME FOR
HAYNES, A "HIGH-ABILITY" STUDENT
IN TREATMENT FIVE

$$\text{Log } Y = \begin{cases} 1. \text{Constant} + b_2(\text{LNPRE}) + b_3(\text{LGTACN}) + b_4(\text{LGTPAR}) + b_5(\text{SEX}) + b_6(\text{SCHL}) \\ 2. \quad 4.9611 + (-0.1041)(1.098) + (-0.2319)(0.51) + (-0.1866)(0.93) + (-0.1419)(0.0) + (0.0872)(1.0) \\ 3. \quad 4.9611 - 0.1143 - 0.1183 - 0.1735 - 0.0 + 0.0872 \end{cases}$$

$$\begin{cases} 1. + b_7(\text{T1}) + b_8(\text{T2}) + b_9(\text{T3}) + b_{10}(\text{T4}) + b_{11}(\text{T6}) + b_{12}(\text{T7}) + b_{13}(\text{T8}) \\ 2. + (-0.0447)(0) + (0.1177)(0) + (0.0877)(0) + (-0.0434)(0) + (0.1643)(0) + (0.1150)(0) + (0.1058)(0) \\ 3. + 0.0 + 0.0 + 0.0 + 0.0 + 0.0 + 0.0 + 0.0 \end{cases}$$

$$\begin{cases} 1. + b_{14}(\text{T9}) \\ 2. + (0.3286)(0) \\ 3. + 0.0 \end{cases}$$

$$\text{Log } Y = 4.6422$$

$$Y = \exp(4.6422) = e^{4.6422}$$

$Y = 104$ minutes, approximately, ESTIMATED TIME FOR HAYNES

For comparison, imagine now a student similarly situated but whose prior achievement scores are one standard deviation below the sample means, and whose matrix algebra pretest score is zero. How much longer will it take this "low-ability" student, Lester, to attain the same criterion performance levels as Haynes? The answer is found by substituting new values where appropriate, as in Table 4. We obtain an estimated time of 204.4 minutes for Lester. Thus equal final attainment can be had for both low and high ability students, but to do so the low student needs almost twice the time spent by the high student. If the cost of time in this technology were constant, then it would be almost twice as expensive to obtain the given level of performance from the low ability student as from the high. It will be shown, however, that in all probability the additional time required by the low student is also more costly per unit.

Table 5

METHOD OF CALCULATING ESTIMATED TIME FOR
LESTER, A "LOW-ABILITY" STUDENT
IN TREATMENT FIVE

$$\text{Log } Y = \begin{cases} 1. \text{Constant} + b_2(\text{LNPRE}) + b_3(\text{LGTACN}) + b_4(\text{LGTPAR}) + b_5(\text{SEX}) + b_6(\text{SCHL}) \\ 2. \quad 4.9611 + (-0.1041)(0.0) + (-0.2319)(-.73) + (-0.1866)(-.55) + (-0.1419)(0) + (0.0872)(1.0) \\ 3. \quad 4.9611 + 0.0 + 0.1693 + 0.1026 + 0.0 + 0.0872 \end{cases}$$

$$\begin{cases} 1. \quad + b_7(\text{T1}) \ldots \\ 2. \quad \ldots \quad \text{(SAME AS Table 4)} \\ 3. \quad \ldots \end{cases}$$

$\text{Log } Y = 5.3202$

$Y = \exp(5.3202) = e^{5.3202}$

$Y = 204.4$ minutes, approximately, ESTIMATED TIME FOR LESTER

In the examples of Tables 4 and 5, the sex and school effects would be as follows: A high-scoring girl would take about 14 minutes fewer than the high-ability boy; the low-scoring girl would finish about 27 minutes sooner than her male counterpart. That is, the girls would take about 13% less time than the boys to reach the same performance level. Both the high and low-scoring boys would take about 9% less time in the second school.

What is the effect of performance level assignment on total time to completion? To answer this question for a student of given characteristics, we need only repeat the procedure used in Tables 4 and 5, substituting new values where required, including the appropriate non-zero treatment dummy. This time an "average" male student will be assumed: one whose pretest, arithmetic concepts and paragraph meaning scores are each at the sample mean. If the treatment dummies all had statistically significant slope coefficients, any such coefficient could be used to represent the effect on required time of the corresponding performance level sequence. Since few of the dummies have consistently significant coefficients, the example here will compare only Treatments 1, 5, and 9, in which students were assigned to the low, middle, or high performance levels throughout the entire sequence. Thus the values to be substituted in the estimated equation are the mean values of the LNPRE, LGTACN, and LGTPAR variables (0.6886, -0.1370, 0.2364, respectively) and the appropriate dummy variable coefficients.

The resulting estimated times for the "average" male student in the first school are:

T1 (65% performance level): 137 minutes

T5 (80% performance level): 143 minutes

T9 (95% performance level): 199 minutes

Thus there is little difference in impact on time between the low and middle performance level sequences, but the time required in the high level sequence is about 40% greater than for the middle sequence.

The time spent by above and below-average-ability male students in the middle performance level sequence was presented in Tables 4 and 5. Predicted times for such students in the low and high performance sequences have also been calculated, and are combined in Table 6 with the below results for an average student.

(INSERT TABLE 6)

Table 6

PREDICTED TIMES IN THREE CRITERION PERFORMANCE LEVEL SEQUENCES BY LOW, AVERAGE, AND HIGH ABILITY STUDENTS

(T1) LOW: 195m. AVG: 137m. HIGH: 99m.	(T2)	(T3)
(T4)	(T5) LOW: 204m. AVG: 143m. HIGH: 104m.	(T6)
(T7)	(T8)	(T9) LOW: 284m. AVG: 199m. HIGH: 144m.

Perhaps the most striking finding to emerge from Table 6 is that a high-ability student can complete the three-lesson sequence at the highest (95%) performance level in less time than the low-ability student takes to complete the sequence at the lowest (65%) performance level. The "average" student takes only slightly longer than the low-scoring pupil in the same comparison. Thus the table dramatically illustrates the importance of allocation of time for the achievement of specific skills by persons of different ability. In this context, the allocation of time is equivalent to the allocation of school and student resources. In particular, schools seeking uniform criterion performance among students (of the kind studied here) in matrix arithmetic, using the technology employed here, would have to allocate space, time with materials, and "teacher" time to low-ability students at about twice the rate for high-ability students.

Further Regression Results:

Interactions. One purpose of the 3 x 3 factorial design of this experiment was to allow statistical testing of whether the sequence of performance level assignments has a significant effect on time to criterion, apart from the magnitude of the performance levels themselves. In the language of analysis of variance, such sequence effects would be called interactions. No significant interactions were found.

The procedure for investigation of interactions was not conventional ANOVA, but rather an exactly equivalent procedure by means of linear regression. In the regression approach to analysis of variance, dummy variables represent main effects and interactions, and significance may be

tested by means of the F-test for addition to explained variance. The steps to be followed are these:

1. Regress covariates alone;
2. Regress covariates and dummy variables for main effects;
3. Regress covariates and dummy variables for treatment cells.

Then an F-test for addition to R^2 from Step 1 to Step 2 tests the significance of main effects. Likewise, the increase from Step 2 to Step 3 is tested for significance of interactions. These results are summarized in Table 7.

Table 7

TESTS FOR SIGNIFICANCE OF MAIN EFFECTS AND INTERACTIONS

Regression	R^2	F-test for significance of addition to R^2
1. Covariates only	0.5734	27.15 (5,101) $p < .005$
2. Covariates plus main effects	0.6279	3.52 (4,96) $p \cong .01$
3. Covariates plus all treatment cells	0.6467	1.22 (4,92) $p > .05$

What does lack of significant interaction mean? It does not mean that no interactions whatsoever are present. It does mean that the data reveal no interactions at a conventional level of significance, in this case the .05 level. It is possible that observation over a longer sequence of in-

structional units, or for different criterion performance levels, would reveal effects pertaining to performance level sequence. An example of such a sequence effect, or interaction, would be if students assigned to the highest performance level for the first two units of instruction took significantly less (or more) time to criterion in the third lesson than did students with lower prior performance level assignments. In the present experiment, however, such effects were not detected.

IV

D I S C U S S I O N

In this experiment, about two-thirds of the variation in total study and testing time spent by students was explained by student characteristic and performance level assignment variables. The estimated production function was used to project the time needed by students of various initial resource combinations to achieve prescribed levels of mastery.

If the cost of time were uniform, it would be twice as expensive to obtain a given level of mastery by a below-average student as by an above-average student.[1] In fact, the cost of time was not uniform, as the lower-ability students required more frequent and prolonged assistance with various aspects of the instructional material than did the superior students. Thus the slower youngsters were also more expensive per minute to teach.

Most studies conclude that yet more research is needed. I will not let the triteness of such a conclusion deter me from asserting it here. The findings of this study strongly suggest that under appropriate conditions the allocation of student (and teacher) time can be a school policy variable of decisive importance. This study suggests that the goal of equity in achievement in specific subject areas or skills will require massive reallocations of school resources. Suppose, for example, a "right to read" program were to be rigorously undertaken. What would an effec-

---

[1] Here, above- and below-averages refer to SAT and pretest scores one standard deviation above and one standard deviation below the mean of students in this study.

tive program cost? The answer, I think, would lie in the technology to be used, the distribution of client entry resources (prior achievement, etc.), and in the minimum performance level adopted. It is quite possible, of course, that the minimum performance level adopted would itself depend on the relative cost of achieving various performance levels. Thus, if there was not a willingness to "pay any price" for literacy, the minimal level chosen would be such as to balance feasible literacy rates against what the public was willing to pay for improved literacy.

The data presented here are limited in several ways: only one technology and one subject matter were used; affective issues were ignored; the experiment was of relatively short duration; the students represent a narrow range of home backgrounds. I suggest, however, that many of these "limitations" are in fact strengths because they are known and because they lend precision to the findings. Nevertheless, many additional issues must be explored before the usefulness of this approach can be determined. The inclusion of students of contrasting socio-economic backgrounds should be a matter of high priority on a research agenda. Likewise, the effect of different technologies and subject matters on the parameters found here should be investigated. We certainly want to know what happens over greater periods of time: Does the bright child's advantage in rate of learning persist, so that the academically rich grow richer, while the poor grow (relatively) poorer? Perhaps the bright youngster's learning is asymptotic in specific situations, so that his advantage lies primarily in the greater variety of subjects he can master while the below-average child is working to master just one. Finally, other age groups must be studied. These

and other matters for investigation should yield findings of use in the analysis and planning of educational policy.